

STATEWIDE CLIENT INDEX USER GROUP MEETING

September 27, 2001 Minutes

Attendees:

Adge, Jeff	DSS-IT
Ajise, Angela	DHS-MEB
Arevalo, Benita	EDS-HFP
Begic, Seija	C-IV
Bornstein, Marty	DHS ITSD
Brill, Roger	DHS ITSD
Calate, Jennifer	CDSS-Fraud
Cisneros, Shirlee	DHS ITSD
Clendenin, Cindy	LEADER
Fong, Michelle	DHS-CMS
Hanna, Waltina	HHSDC-SFIS
Hartz, Barbara	DHS MCH
Hill, Alicia	ISAWS
Ho, Le-Ba	CaWIN
Kick, Brian	WDTIP
Lara, Elena	DHS MEB
McCarley, Traci	DHS-CMS
McCreary, Maureen	DHS-MEB
Olson, Pete	DHS ITSD
Oxford, Sherena	C-IV
Parman, Debbie	Nevada County
Perry, Sharon	DHS ITSD
Pierce, Rick	Stanislaus County
Preece, Mary	Ventura County
Sabel, Kathy	DHS ITSD
Sheldon, Meg	CWDA
Tabor, Karen	DHS-WIC
Tirado, Maria	ISAWS
Wang, Jina	DHS ITSD

Purpose of the meeting

- To get user feedback regarding whether any of the proposed potential enhancements would improve the accuracy of the CIN assignment and/or improve the efficiency of the folks performing file clearance; and if so, to help us set priorities.
- To discuss issues and processes that need resolution in order to deal with existing data problems.

Transaction Formats

Issue: Since changing the SCI record layout format would require that end users make corresponding changes to their systems, and not all users may wish to send/receive the additional data, then we are constrained from adding new functionality for address-based matching or for family-based matching. The 174 bytes allocated for the input details have very little FILLER left to add new data; similarly the 147-byte chunks of data used in sending back multiple potential matches for a query also has only a few bytes of unused FILLER.

Recommended solution: Make use of the existing version control field on the input header

Analysis:

All of the existing transactions would be version 01. If we develop any new formats for future transactions, we could assign these to be version 02 types. So **if the first 2 bytes of the input header were either blank or '01', then we would expect input transactions in the current format and we would return output in the current format.** However if that field contained '02', then we would expect the input transaction to be in the new format and would know that the end user is prepared to receive output in the new format.

The advantage of this approach is that applications with an existing SCI interface would not be required to make any changes to their systems. The disadvantage is that, under certain circumstances, the State might need to maintain two versions of a program. This seems to be a pretty good trade-off for us.

Feedback from the group: This seems to be a reasonable solution.

SCI hours

A long term goal is for SCI to be a 24 x 7 system, but at this point, no users have defined this to be a business requirement, so round-the-clock availability is not a high priority.

We are planning to make a change in October so that the SCI hours of availability would be consistent with MEDS.

That means that the online system would shut down at midnight and be brought back up at 2 AM. Currently it is shut down at 2 AM and brought back up again at 4 AM. This will minimize the likelihood of either of two things occurring:

1. Certain MEDS transactions, such as the EW15, EW20, and planned transactions for a new program, the Breast and Cervical Cancer Treatment program, failing because they could not access SCI.

2. The online SCI OM transaction, which gets MEDS data, failing because MEDS is down.

If a user needs SCI to be available to available between midnight and 4 AM (for example, if a county needed to work extended hours in order to convert to CalWIN), then it is very possible that we could extend the SCI hours, at least temporarily.

Feedback from the group: This change will not cause any adverse actions at this time.

NOTE: This change of hours has now been installed. The SCI training region hours of operation were also changed to also be the same as the SCI production hours.

Concept for a Family Based search

One action item from the April 2001 meeting was to explore the feasibility of adding functionality to SCI so that end users could perform a family-based match. The purpose would be to assist the end user to determine if an existing record belongs to your client. The most common time when this function might be useful is when it appears that there might be multiple records on file that may belong to your client. It might also be useful if there was a single potential match that was close, but not close enough for you to be completely comfortable to link to this existing CIN.

There is a presumption that if the CIN that you think may belong to your client is on a MEDS record whose other case members correspond with your client's family, then it is a pretty safe bet that this is the right CIN and visa versa. **This function would not provide better information than if you were to perform whole case look-ups on MEDS, but it could be designed to save user time and effort by having the process automatically perform a series of steps.**

To understand this process, it is useful to know how data is arrayed on MEDS.

- The MEDS data base record contains a complete set of data about each individual.
- But to determine which persons are members of that person's case, one would need to read a cross reference file, called the County ID cross reference file. This particular file has a very limited set of information - the case number, the SSN (or MEDS ID), and the birthdate. It does not contain the names of the family members.

In both alternative solutions, the front end would be the same. We would develop a new SCI transaction. The user would need to provide either an SSN or a CIN. When SCI receives this transaction, it would read the MEDS cross-reference files to get a list of the County IDs associated with this client.

NOTE: ITSD staff did not perform an actual frequency distribution of how many cases that individuals are known by, but instead did a visual scan of random portions of the MEDS County ID cross reference file. There are some clients who are known by more than 20 County IDs, but most are known by 1 to 6 County IDs. For the subset, which we analyzed, the average seemed to be around 3.

If there are no County IDs associated with this client, we can not do a family-based search. SCI would return an error code.

If there were only one occurrence, then SCI would return a set of data. *Feedback from the group: this proposed transaction would need to return at least the name, birthdate, and MEDS ID/SSN of all case members. The requirement for returning the name means that this process must read the MEDS data base record for each case member.*

Where the two approaches differ is when there are multiple County IDs found for this client.

One alternative would be for SCI to return a list of County IDs to the end user. We would design some logic to screen out the "phony" County IDs, which have either the client's SSN or CIN rather than a county-assigned CIN. These clients could have been reported to MEDS by the federal government, Healthy Families, or a health program like CMS Net. Screening out the phony case IDs will prevent a false positive match that could occur if the first 7 digits of someone's SSN or CIN happen to match a county Case number. The user would send a second transaction with the selected county ID and then SCI would return the appropriate set of data.

The other approach would be for SCI to select the most likely County ID based on several factors, which could include the following:

- looking at the dates on the County IDs cross reference file to determine which are the most recently reported values;
- using a hierarchy of County IDs, for example looking at first at pending data, then current Medi-Cal, then Food Stamps, then special programs,
- determining which County IDs have aid codes, which indicate that there are likely to be multiple persons in the case (for example, those that begin with '3'). SCI would return the set of data for that County ID.
- Using the county code of the user as a selection criteria.

We would screen out inappropriate County IDs.

Feedback from the group: there is a presumption that end users would have searched their local system first so that the presence of a County ID from your county should not be given a great weight in this proposal.

It would have been feasible to return data for all County IDs if the response was limited to MEDS IDs/SSNs and birthdates, but to get the family member names we would need to do separate reads of MEDS. Because there can be a large number of County IDs, a large number of people in the case, or both, that means there is a potential that the process could require so many reads that it would not only result in a very poor response time for that specific transaction, but could also degrade the response time for all users.

Feedback from the group:

The approach described above, where the system selects the "best" County ID for each client, is the preferred solution. The users recognize that, under this approach, they would not be able to use SCI to perform a family based match on other County IDs known to this client. If the family data for the best County ID did not match, then the users could, if they wish, perform a family-based search for the next potential match on the pick list. The State will send out a proposal about what rules would be used to select the best County ID. The enhancement will be designed based on feedback from the users.

Collecting middle name from DMV records

Having more records with middle names should make file clearance more accurate, particularly for processing common names. Currently around 30% of SCI records have data in the middle name field. There are two factors that will be likely to increase the number of records on SCI with middle names.

1. One action items from the April meeting was to follow up with SSA to find out if the presence of middle names or middle initials would impact the number of records that pass SSA's SSN validation edits, either positively or negatively. SSA responded that the impact would be expected to be positive, in other words, increase the number of records that pass validation.
This information has been forwarded to DMV and they are currently analyzing whether or not to modify their process to include middle name data on their SSN Validation queries. If DMV does make this change, then this could serve to increase the number of SCI records with middle name information, since DHS has an agreement with DMV that we can populate our files with data from their records that validate and are known to our system.
2. The Health Insurance Portability and Accountability Act (HIPAA) requires the presence of the middle initials for clients who will be enrolling or disenrolling in managed care plans. Approximately 70% of current clients on MEDS have been enrolled in managed care, so over time, we would expect that the number of SCI records, with data in the middle name field, will increase greatly. As DHS implements new programs, such as the Breast and Cervical Cancer Treatment program, middle name will be a required field on those applications. The planned install date for this portion of HIPAA is October 2002.
NOTE: Congress has subsequently voted to delay implementation by 12 months.

We think that it would be helpful to modify SCI to include the middle name in the responses to alternate name queries. The only customers utilizing the SCI Alternate Name query function are ISAWS, LEADER, and WDTIP. We will send a note to these applications asking if they would like us to make this change and if so, when.

We found a pretty sizable number of records with low values in the middle name field. Since some programs treat the presence of low values as an indicator that the record has no other data in it, we will change these records to have spaces in the middle name field. We will also research why we are getting low values in this field and make the appropriate changes to SCI.

NOTE: subsequent analysis of the SCI logic found that low values were identified and fixed in the batch add and interactive update process, but not in the interactive add process. The coding to correct this problem was installed on October 26th.

Address searches

In the April 2001 meeting, the consensus of the group was that if we were to incorporate an address search in SCI, then it would probably be better to use zip code or zip + 4 or zip + 4 + 2. It was suggested that users perform some type of field testing to compare how often the address on the application matched the address on MEDS. LA County staff performed these tests and found that approximately half of the application addresses matched the MEDS addresses. The clients, who were not previously known to MEDS, were included in the portion that did not match.

Since April, we have gained experience with interactive address matching using bar codes, so we want to also put this alternative on the table. We could set up a second version of the scored search, which would allow the customer to input the client's address on the input transaction. We would convert that address to a barcode and match the bar-coded-address to the address on MEDS. It would be difficult to assign a percentage score because that scoring module is performed done within the SSA Name3, while this bar-coded address matching would be external to SSA Name3

So the response would be an indicator that either 'Yes, there is a match'; 'No, the address did not match'; or 'Unknown, we did not have a previous record on file for this client so we couldn't do a match.'

If instead we use variations of the zip code as search parameters, we'd probably incorporate this into a new version of the SCI 'OI' transaction. Advantages of using zip code include:

1. We could store multiple occurrences without using a great deal of disk space as we would if we contemplated storing multiple occurrences of the entire address. By storing multiple occurrences, we mitigate the problems associated with the fact that clients do tend to move around.
2. Returning the zip code in the datastream does not have the same problems of confidentiality as returning the full address.
3. Although it would be tricky to add this to the scored search, it is doable. It is tricky because addresses are dynamic. A match is a much better indicator that this is the right person than a mismatch is an indicator that this is not the same person.

In addition, we could incorporate some address-based logic in the batch file load procedure before we begin on the CalWIN rollout, which would make those matches more accurate.

Feedback from the group:

- *Storing multiple occurrences of the zip code would not add much value.*
- *If we were ever to install this match, we should not use a mismatch of an address to screen out a potential match for an interactive query. Instead we should simply use the presence of a match or mismatch as a factor in determining the order in which we should return the pick list of potential matches.*
- *The mobility of the population and the potential confusion between mailing address and residence address greatly limit the usefulness of an address match. This potential enhancement will be assigned a low priority, which means that unless we get different feedback in the future, it is unlikely that this function will be incorporated into SCI.*

MEDS data via SCI

DHS is in the process of adding new data elements to the SCI transaction to obtain MEDS data, as requested by CalWIN. These new data elements, MEDS name, the SSA birthdate, and the birthdate validation code, will be added at the end of the existing data stream. This will allow current users of this transaction, SFIS and LEADER, to opt to modify their programs to obtain the new data or to leave their programs as is and simply treat the new output fields as Filler.

We have completed unit testing and acceptance testing. We are working with end users to perform system testing.

NOTE: This change was installed into production on October 12, 2001.

Changing the Way that MEDS Collects CINs

In the April meeting, we talked about the issue of MEDS changing the way it collects CINs. The purpose was to handle those cases where the application knew the client by a different CIN than MEDS does. Phase I of this change was installed in late August 2001. Now MEDS will typically create a cross-reference pointer to the other CIN.

There are two related issues to discuss.

Since SCI will show that this second CIN is known to MEDS, should we change the Known-to-MEDS flag values?

Now spaces means that it is not known to MEDS and 'Y' means it is known. There are different alternative solutions.

- A) no change....'Y' will mean that the CIN is known to MEDS, but it won't now mean that this is CIN by which MEDS knows this client. Because it also might mean that this is the CIN that was reported to MEDS by another application.
- B) We could define 'Y' to mean that this is the CIN by which MEDS knows the client and have another value for secondary or other (for example, 'O').

Feedback from the users: It would be a good thing to report which CIN is the primary CIN, but not until end users are prepared to accept new values for known-to-MEDS flag.

A related issue is whether we should combine the two SCI records under certain circumstances.

NOTE: we did not have enough time to discuss this particular issue when we had the discussion on handling duplicate records.

Forced Link proposal

In order to minimize the likelihood that the same CIN will be used for two different clients, we are contemplating not allowing links when the data appears to be mismatched. This discussion has two components – how do we determine if the data is mismatched and secondly if it is mismatched, what options does the end user have.

The easiest way to identify a potential mismatch is to add an edit to check if at least two of these conditions are true:

- Link trans first/last names differs from SCI current or all previously reported names
- Link trans birthdate differs from SCI current or all previously reported birthdates
- Link trans gender differs from SCI gender
- Link trans SSN numeric, differs from SCI current or all previously reported numeric SSNs

Another alternative, which would take more effort, but is likely to provide more accurate results, would be for us to incorporate a scoring module, similar to the name search scoring logic, inside of the link function. If the score was below a certain threshold then we wouldn't do the link. This would allow end users to perform a link, for example, when there was a transposition in the SSN and the client is now using her married name.

End users would have three alternatives if we incorporated this logic:

- The alternative, which takes the least effort, would be to select a new CIN.
In most, but not all, cases this would be the best choice in terms of data integrity.
- If the user wanted to use this CIN for this client, then he or she could send an e-mail to the SCI Database Administrator via the ITSD Help Desk to manually link this record.

- Another solution, which will allow the end user to link these records, but without making as much effort would be for SCI to develop a Forced Link transaction, which is similar in concept to the Forced Add transaction. In other words, it would skip those edits that look for mismatches and just do the link. Obviously this would require end users to change their screens and their automated interface programs.

Feedback from the group:

- *SCI should not add a new filter to screen out potential mismatches at this time. Both LEADER and CaWIN plan to incorporate logic so that end users will need to validate that any demographic changes to an existing record are appropriate. Thus there is a presumption that the end user (in these systems) would have already made the determination that the updating of multiple fields is appropriate.*
- *DHS should count the number of occurrences of records where multiple fields are being updated. If the number is larger than expected, then we may want to hold further discussions on this proposal.*
- *SCI should capture the SSN Validation code and then add edits, which would not allow updates to certain fields if the record has been validated. For example, let's say that the SSN has been validated and SCI received a link or an update transaction with a different SSN. Then that different SSN would not be posted to what SCI considers the best set of data for the client (called the SCI Main Table), but instead would get posted to the file which contains all of the SSNs by which this client has been known. So when someone does a standard query for this client, the validated SSN would be returned. To see the other SSN, a user would need to perform an Alternate SSN query.*

Change inquiry process

The Healthy Families Program (HFP) requested that we explore the possibility of making changes to the SCI search process flow. Currently SCI will automatically perform a name search (if the user supplied the client's name, birthdate, and gender) if the input SSN is not found on the SCI tables. However if an SSN match is found, then SCI will not do a name match. What HFP would like to explore is the possibility of performing name matches even if the SSN is found.

We would need to have the user enter a search type of the transaction.

We would either define a new search type of 'D' (dual search).

1. In either case, the new process flow would be:
 1. If search type = 'D', perform standard name search, even if SSN > spaces. Put results in a table.
2. If there are not 25 potential matches above the minimum threshold for returning interactive data, then perform search on the SSN, and append potential matches at the end of the table (until the table is full or the SSN-based matches have been included, whichever comes first).
3. Unduplicate by CIN.

NOTE: if HFP or one of public health programs, using the CATS hub to access SCI, wanted to use this function, we would need to establish a new CATS transaction, perhaps a 0101 trans.

Feedback from users:

- *To eliminate the likelihood that a match on SSN might not be returned for common names, reverse the proposed process, so that the SSN match is done first, then do the name match.*
- *Users other than HFP would also be interested in using this function if it becomes available.*

Trouble-shooting procedures

A copy of the most current recommended trouble-shooting procedures for handling connectivity, data, or FTP problems will be sent out with the User Group notes. We will also send a copy to folks on the distribution list to be notified if there is a planned or unplanned system outage.

Web Page

We are proceeding to develop a web page for SCI. The home page is designed. The internal links have been created. We are still developing the content. The hang-up is to the link with the set of best practices. Because everyone does business somewhat differently, we may include multiple sets of best practices.

Action items:

1. ITSD will notify end users when the operating hours are changed so that SCI is brought down at midnight **(DONE)**.
2. ITSD will send out a proposal, requesting feedback from all users, regarding what rules would be used to select the best County ID for a family based search.
3. ITSD will send out a proposal, requesting feedback from ISAWS, LEADER, and, WDTIP, about adding the middle name to the Alternate Name query response.
4. ITSD will notify LEADER, SFIS, and CalWIN when the new version of the SCI transaction to obtain MEDS data is implemented **(DONE)**.
5. ITSD will plan to modify SCI to be able to differentiate which CIN is the primary CIN for MEDS when a client is known by more than one CIN on MEDS. However, until all of the end users are ready to accept the new value in the known-to-MEDS field, SCI will continue to return the value 'Y' if the client is known to MEDS, regardless of whether this is the primary or the secondary CIN on MEDS for this client.
6. ITSD will count the number of occurrences whether there is an update or a link, which changes two or more of these fields (SSN, but only if the existing SSN is numeric; name; birthdate; or gender) on the existing records for a client.
NOTE: this is being **done** via an attachment to these user group notes.
7. ITSD will raise the priority level for adding the SSN Validation Code to the SCI tables. We will send out a proposal to all users, which will contain the following recommendations:
 - a) will SCI be updated with the MEDS SSN validation code if the SCI name and birthdate do not match the MEDS name and birthdate?
 - b) after SCI has captured this data, should certain potential updates to the SSN, name, or birthdate fields be captured on the alternate tables, but not on the main table (and thus would not be displayed as part of a standard query response)?
8. ITSD will send a note to all users, who do not currently receive a referral file, describing the available options for receiving notifications of demographic changes to those records where the CIN is associated with your clients.
NOTE: this is being **done** via an attachment to these user group notes.
9. ITSD will make the appropriate changes to SCI and CATS to offer users the ability to perform a combined SSN and name search.

10. ITSD will send out a description of the SCI delete function and will tell end users how they could access this function, if they so wish.

NOTE: this is being **done** via an attachment to these user group notes.

11. After the revised EW11/GZ11 process is finalized, we will set up another User Group meeting with an emphasis on resolving the major issues for handling duplications and mislinks.
12. ITSD will correct those records, which have low values in the Middle name field and will research why these occurred so that we can install some preventive measures.

NOTE: the **design changes have been done** so no new records should be created with low values in any of the fields. The **clean-up of the existing records was done** in the last weekend of October. There were slightly less than 250,000 records with low values on the SCI tables. All occurrences of low were either in the Middle Name field or the Alien Number field.